

# THE EXIT PROBLEM IN OPTIMAL NON-CAUSAL ESTIMATION

D. Ezri\*      B. Z. Bobrovsky†      Z. Schuss‡

February 5, 2008

## Abstract

We study the phenomenon of loss of lock in the optimal non-causal phase estimation problem, a benchmark problem in nonlinear estimation. Our method is based on the computation of the asymptotic distribution of the optimal estimation error in case the number of trajectories in the optimization problem is finite. The computation is based directly on the minimum noise energy optimality criterion rather than on state equations of the error, as is the usual case in the literature. The results include an asymptotic computation of the mean time to lose lock (MTLL) in the optimal smoother. We show that the MTLL in the first and second order smoothers is significantly longer than that in the causal extended Kalman filter.

**Keywords:** Nonlinear smoothing, loss of lock, cycle slips

## 1 Introduction

In many applications in communication practice a random signal  $\mathbf{x}(t)$  is received through a noisy channel. The random signal  $\mathbf{x}(t) \in \mathbb{R}^N$  is assumed to be a stochastic

---

\*Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel. email: ezri@eng.tau.ac.il

†Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel. email: bobrov@eng.tau.ac.il

‡Department of Mathematics, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel. email: schuss@post.tau.ac.il

process defined by an Itô stochastic differential equation (SDE) [17]

$$d\mathbf{x} = \mathbf{m}(\mathbf{x}, t) dt + \boldsymbol{\sigma}(\mathbf{x}, t) d\mathbf{w}, \quad (1)$$

where  $\mathbf{w}(t)$  is a vector of standard Brownian motions. The measurements process  $\mathbf{y}(t) \in \mathbb{R}^M$ , which is the output of the noisy channel, is modeled by another SDE

$$d\mathbf{y} = \mathbf{g}(\mathbf{x}, t) dt + \sqrt{N_0/2} d\mathbf{v}, \quad (2)$$

where  $N_0$  measures the channel noise intensity and  $\mathbf{v}(t)$  is another vector of Brownian motions, independent of  $\mathbf{w}(t)$ . We further assume that the functions  $\mathbf{m}(\cdot, \cdot)$ ,  $\boldsymbol{\sigma}(\cdot, \cdot)$  and  $\mathbf{g}(\cdot, \cdot)$  satisfy standard regularity conditions such that (1),(2) possess a strong unique solution.

When the optimality criterion is minimum square error, the optimal filtering problem is to construct the causal estimator  $\hat{\mathbf{x}}(t) = E[\mathbf{x}(t) | \mathbf{y}(s)]$  of  $\mathbf{x}(t)$ , where  $0 \leq s \leq t$  [32]. The optimal fixed interval smoothing problem is to construct the non-causal estimator  $\hat{\mathbf{x}}(t) = E[\mathbf{x}(t) | \mathbf{y}(s), 0 \leq s \leq T]$ , where  $T$  is the length of the interval, and  $t < T$ . The optimal fixed lag smoothing problem is to construct the non-causal estimator  $\hat{\mathbf{x}}(t) = E[\mathbf{x}(t) | \mathbf{y}(s), 0 \leq s \leq t + \Delta]$ , where  $T$  is the length of the interval, and  $t + \Delta < T$ . In many applications delay in the estimation of the signal is not permissible, as for example in closed-loop control, radar tracking systems, and so on. There are, however, interesting cases, where certain delay is permissible, as for example in communication systems, as extensively practiced in coding [34, 20].

Smoothers are used because their performance is superior to all causal filters, with respect to the same optimality criterion [19]. In linear estimation theory, where the optimality criterion is minimum mean square error, the error variance of the optimal smoother is smaller than that of the optimal filter [10].

Optimal estimators are usually infinite-dimensional and therefore have no finite-dimensional realizations, so that suboptimal estimators have been proposed to approximate the optimal ones by a system of SDEs, driven by the measurements [13, 21, 22]. The phase-locked-loop (PLL), which is a realization of the extended Kalman filter (EKF) [30], is a nonlinear suboptimal causal estimator of the carrier phase in various communications systems [16]. A well known effect in such PLL demodulators is the cycle slip phenomenon that consists in occasional sudden changes of size  $2\pi n$  ( $n = \pm 1, \pm 2, \dots$ ) in the phase estimation error [5]. Obviously, the mean time between cycle slips, known as the mean time to lose lock (MTLL), decreases with the noise intensity and causes sharp degradation in the performance of the filter and to the formation of a performance threshold [33, 29, 28].

Considerable effort was put into the computation of the MTLL in causal estimators [16, 24, 31, 33], including the singular perturbation method [28, 29] and large deviations theory [9, 8]. However, the phenomenon of loss of lock in smoothers has never been addressed, despite the extensive study of linear and nonlinear smoothers in the literature [21, 22, 23, 14, 35, 6, 2, 25, 26, 10, 15, 1]. The objective of this paper is to provide the missing theory, estimate the MTLL in the optimal smoother, and compare it with that in the casual PLL. Specifically, we compute the asymptotic distribution of the optimal estimation error in case the number of trajectories in the

optimization problem is finite. We identify the contribution of error trajectories to the minimum noise energy (MNE) cost functional and recast the problem in terms of order statistics. The asymptotic expression for the MTLL in the smoother is similar to that resulting from the Wentzell-Freidlin theorem for causal systems, with a new functional. Applying our method to standard phase models, we show that the MTLL in the optimal smoother is significantly longer than that in the PLL.

## 2 The mathematical model

The general equations of a scaled phase tracking system consist of the linear model of the phase  $\mathbf{x}(t) = [x(t), x_2(t), \dots, x_N(t)]^T$  [29]

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \sqrt{\varepsilon} \mathbf{B} \dot{\mathbf{w}} \quad (3)$$

and the nonlinear model of the noisy measurements  $\mathbf{y}(t) = [y_s(t), y_c(t)]^T$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \sqrt{\varepsilon} \dot{\mathbf{v}}, \quad (4)$$

with

$$\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \sin x \\ \cos x \end{pmatrix}.$$

The dimensionless parameter  $\varepsilon$  is assumed small in the case of a low noise channel [28].

A fixed-interval minimum noise energy (MNE) estimator  $\hat{\mathbf{x}}(\cdot)$  for  $\mathbf{x}(\cdot)$  is the minimizer of the cost functional [6]

$$J[\mathbf{z}(\cdot)] \equiv \int_0^T [|\mathbf{y} - \mathbf{h}(\mathbf{z})|^2 + |\dot{\boldsymbol{\zeta}}|^2] dt, \quad (5)$$

with the equality constraint

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}\dot{\boldsymbol{\zeta}}, \quad (6)$$

that is,

$$\hat{\mathbf{x}}(\cdot) \equiv \arg \min_{\mathbf{z}(\cdot)} J[\mathbf{z}(\cdot)]. \quad (7)$$

Note that the integral  $J[\mathbf{x}(\cdot)]$  contains the white noises  $\dot{\mathbf{w}}(t)$ ,  $\dot{\mathbf{v}}(t)$ , which are not square integrable. To remedy this problem, we begin with a model in which the white noises  $\dot{\mathbf{w}}(t)$ ,  $\dot{\mathbf{v}}(t)$  are replaced with square integrable wide band noises, and at the appropriate stage of the analysis, we take the white noise limit (see below).

In contrast to nonlinear filtering, where the locked state is a local attractor for the error dynamics [29], and escaping it corresponds to loss of lock, there is no dynamics, and therefore no attractors for smoothers. Thus, we have to define cycle slip events in a different manner than hitting the boundary of the domain of attraction. We define the estimation error  $\mathbf{e}(t) = [e(t), e_2(t), \dots, e_N(t)]^T$  as

$$\mathbf{e}(t) = \hat{\mathbf{x}}(t) - \mathbf{x}(t), \quad (8)$$

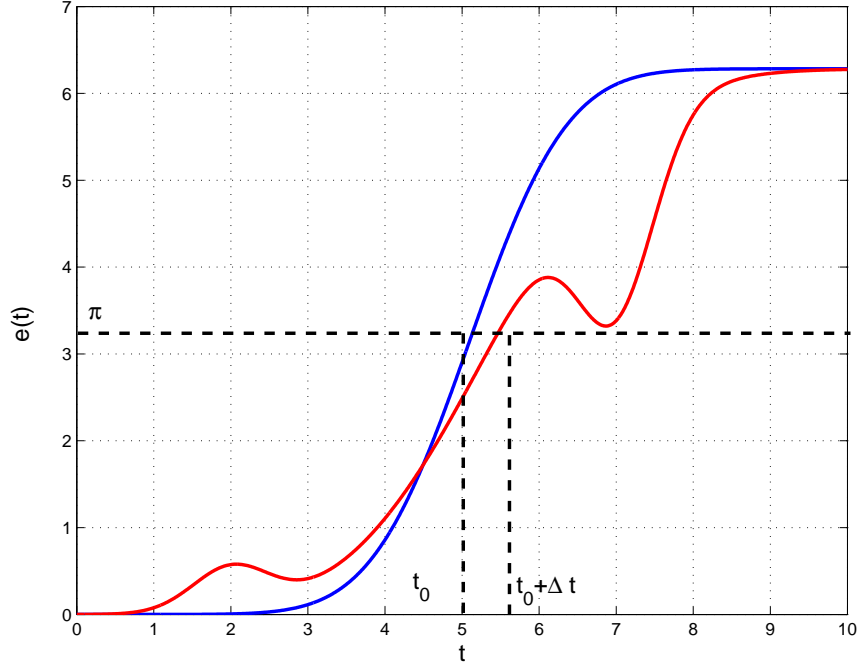


Figure 1: An example of two trajectories in  $\mathcal{C}^1(t_0 = 5)$

and we say that a cycle-slip has occurred in the time interval  $(t_0, t_0 + \Delta t)$ ,  $\Delta t \ll 1$ , if and only if the estimation error  $e(t)$  vanishes at at some  $t_0 - T_1$ , reaches the point  $[2\pi n, \mathbf{0}]^T$ , ( $n = \pm 1, \pm 2, \dots$ ), at some later time  $t_0 + T_2$ , and  $e(\tilde{t}) = \pi n$ , where  $\tilde{t} \in (t_0, t_0 + \Delta t)$ . The time  $T_s = T_1 + T_2$  is the slip duration, satisfying  $T_s \ll T$ . We define in the space of continuous functions  $\mathcal{C}_{[0,T]}^N$  the set  $\mathcal{C}^N(t_0)$  of all continuous trajectories  $e(\cdot)$  that slip in the interval  $(t_0, t_0 + \Delta t)$ . Thus

$$\Pr \{\text{slip in } (t_0, t_0 + \Delta t)\} = \Pr \{e(\cdot) \in \mathcal{C}^N(t_0)\}. \quad (9)$$

For small values of  $\varepsilon$  the cycle-slip event is a rare large deviation from the original trajectory, and therefore  $\hat{\mathbf{x}}(t)$  is in the vicinity of  $\mathbf{x}(t)$  before the cycle-slip, and in the vicinity of  $\hat{\mathbf{x}}(t) + [2\pi n, \mathbf{0}]^T$  after the slip. Thus, we can define the beginning and the end of the cycle-slip by the instants when  $e(t)$  reaches the origin and  $[2\pi n, \mathbf{0}]^T$ , respectively. An example of two trajectories in  $\mathcal{C}^1(t_0 = 5)$  is given in Figure 1.

### 3 The MTLL in the optimal smoother

The Wentzell-Freidlin theorem [8, 9] and the singular perturbation method [29] for asymptotic evaluation of the MTLL are concerned with stochastic processes satisfying a stochastic differential equation with a unique solution. In contrast, the dynamics of the optimal smoother, derived from the EL equations, form a two-point boundary-value problem which has no unique solution. Therefore the Wentzell-Freidlin and the singular perturbation method seem inappropriate for the computation of the MTLL in a smoother. It appears that this computation calls for a different approach.

The first step toward an asymptotic calculation of the MTLL in smoothers is the computation of the asymptotic distribution of the estimation error in case the number of trajectories in the optimization problem is finite. We investigate the cost functional of deterministic error trajectories that deviate from the original trajectory  $\mathbf{x}(t)$ . We augment  $\mathbf{x}(t)$  with the set of the  $N_T$  trajectories  $\mathbf{r}_i(t) = [r_i(t), r_i^{[2]}(t), \dots, r_i^{[N]}(t)] \in \mathcal{C}_{[0,T]}^N$ ,  $i \in [1, \dots, N_T]$ . The trajectories  $\mathbf{x}(t) + \mathbf{r}_i(t)$  are admissible in the optimization problem (5), (6), only if the trajectories  $\mathbf{r}_i(t)$  satisfy

$$\dot{\mathbf{r}}_i = \mathbf{A}\mathbf{r}_i + \mathbf{B}\mathbf{u}_i. \quad (10)$$

We define the difference

$$\Delta J[\mathbf{x}(\cdot), \mathbf{r}_i(\cdot)] \triangleq J[\mathbf{x}(\cdot) + \mathbf{r}_i(\cdot)] - J[\mathbf{x}(\cdot)] \quad (11)$$

and substitute (5), (6) and (10) in (11) to obtain

$$\begin{aligned} \Delta J[\mathbf{x}(\cdot), \mathbf{r}_i(\cdot)] &= \int_0^T \left[ 4 \sin^2 \left( \frac{r_i}{2} \right) + |\mathbf{u}_i|^2 \right] dt \\ &+ \sqrt{4\varepsilon} \int_0^T [\sin x - \sin(x + r_i)] dv_1(t) \\ &+ \sqrt{4\varepsilon} \int_0^T [\cos x - \cos(x + r_i)] dv_2(t) \\ &+ \sqrt{4\varepsilon} \int_0^T \mathbf{u}_i^T d\mathbf{w}(t). \end{aligned} \quad (12)$$

At this point, we take the white noise limit in the wide band noises so the stochastic integrals in (12) become Itô integrals. Collecting them into a single Itô integral leads to

$$\begin{aligned} \Delta J[\mathbf{x}(\cdot), \mathbf{r}_i(\cdot)] &= \int_0^T \left[ 4 \sin^2 \left( \frac{r_i}{2} \right) + |\mathbf{u}_i|^2 \right] dt \\ &+ \sqrt{4\varepsilon} \int_0^T \sqrt{4 \sin^2 \left( \frac{r_i}{2} \right) + |\mathbf{u}_i|^2} d\tilde{v}_i(t), \end{aligned} \quad (13)$$

where  $\tilde{v}_i(t)$  is a standard Brownian motion that depends on  $\mathbf{v}(t)$  and  $\mathbf{w}(t)$ .

Note that although the values of the random variable  $\Delta J[\mathbf{x}(\cdot), \mathbf{r}_i(\cdot)]$  depend on the trajectories of  $\mathbf{x}(t)$  and  $\mathbf{y}(t)$  through  $\mathbf{v}(t)$  and  $\mathbf{w}(t)$ , the probability law of  $\Delta J[\mathbf{x}(\cdot), \mathbf{r}_i(\cdot)]$  depends only on the trajectories  $\mathbf{r}_i(t)$ . Therefore, we abbreviate notation to  $\Delta J[\mathbf{r}_i(\cdot)]$ . We note further that  $\Delta J[\mathbf{r}_i(\cdot)]$  is a Gaussian random variable with expectation

$$\mathbb{E} \Delta J[\mathbf{r}_i(\cdot)] = m_i, \quad (14)$$

where

$$m_i \triangleq \int_0^T \left[ 4 \sin^2 \left( \frac{r_i}{2} \right) + |\mathbf{u}_i|^2 \right] dt, \quad (15)$$

and variance

$$\text{Var}\{\Delta J[\mathbf{r}_i(\cdot)]\} = 4\varepsilon m_i. \quad (16)$$

Furthermore, we compute the covariance

$$\begin{aligned} \sigma_{ij} &= \mathbb{E}(\Delta J[\mathbf{r}_i(\cdot)] - m_i)(\Delta J[\mathbf{r}_j(\cdot)] - m_j) \\ &= 4\varepsilon \int_0^T \mathbf{u}_i^T \mathbf{u}_j dt + 4\varepsilon \int_0^T [1 + \cos(r_i - r_j) - \cos r_i - \cos r_j] dt. \end{aligned} \quad (17)$$

Note that if the supports of  $\mathbf{r}_i(\cdot)$  and  $\mathbf{r}_j(\cdot)$  are disjoint, the cost functionals  $\Delta J[\mathbf{r}_i(\cdot)]$ ,  $\Delta J[\mathbf{r}_j(\cdot)]$  are not correlated. We conclude that the random variables  $\Delta J[\mathbf{r}_i(\cdot)]$ ,  $i = 1 \dots N_T$ , form a Gaussian random vector with distribution

$$\begin{bmatrix} \Delta J[\mathbf{r}_1(\cdot)] \\ \Delta J[\mathbf{r}_2(\cdot)] \\ \vdots \\ \Delta J[\mathbf{r}_{N_T}(\cdot)] \end{bmatrix} \sim N \left\{ \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{N_T} \end{bmatrix} ; 4\varepsilon \begin{bmatrix} m_1 & \sigma_{12} & \dots & \tilde{\sigma}_{1N_T} \\ \tilde{\sigma}_{12} & m_2 & & \\ \vdots & & \ddots & \\ \tilde{\sigma}_{1N_T} & & & m_{N_T} \end{bmatrix} \right\}, \quad (18)$$

where  $\tilde{\sigma}_{ij} = \frac{1}{4\varepsilon} \sigma_{ij}$ .

When considering a finite number of error trajectories  $\mathbf{r}_i(\cdot)$ ,  $i = 1, \dots, N_T$  in the optimization problem. The estimator error trajectory  $\mathbf{e}_{N_T}(\cdot)$  minimizes the cost functional  $\Delta J[\mathbf{r}_j(\cdot)]$ ,

$$\Pr\{\mathbf{e}_{N_T}(\cdot) = \mathbf{r}_k(\cdot)\} = \Pr\{\Delta J[\mathbf{r}_k(\cdot)] < \Delta J[\mathbf{r}_j(\cdot)] \text{ for all } j \neq k\}. \quad (19)$$

Thus the problem of minimization has been recast in the language of order statistics. The probability on the right side of (19) is difficult to calculate, so we pursue the distribution of the error trajectory  $\mathbf{e}_{N_T}(\cdot)$  in the limit of small  $\varepsilon$ . We assume that for each  $k$  there exists an interval  $A_k$  such that

$$\mathbb{E}\{\Delta J[\mathbf{r}_j(\cdot)] \mid \Delta J[\mathbf{r}_k(\cdot)] \in A_k\} > \Delta J[\mathbf{r}_k(\cdot)] \quad \text{for all } j \neq k. \quad (20)$$

Cramér's theorem for Gaussian vectors [8, 7] implies that in the limit of small  $\varepsilon$

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \Pr\{\Delta J[\mathbf{r}_k(\cdot)] < \Delta J[\mathbf{r}_j(\cdot)] \text{ for all } j \neq k\} = \\ \lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \Pr\{\Delta J[\mathbf{r}_k(\cdot)] \in A_k\}. \end{aligned} \quad (21)$$

Next, we evaluate the conditional expectation

$$\mathbb{E}\{\Delta J[\mathbf{r}_j(\cdot)] \mid \Delta J[\mathbf{r}_k(\cdot)]\}. \quad (22)$$

Since  $\Delta J[\mathbf{r}_j(\cdot)]$ ,  $\Delta J[\mathbf{r}_k(\cdot)]$  are jointly Gaussian,

$$\begin{aligned} \mathbb{E} \{ \Delta J[\mathbf{r}_j(\cdot)] \mid \Delta J[\mathbf{r}_k(\cdot)] \} &= m_j + \frac{\sigma_{jk}}{4 \varepsilon m_k} (\Delta J[\mathbf{r}_k(\cdot)] - m_k) \\ &= m_j + \frac{\tilde{\sigma}_{jk}}{m_k} (\Delta J[\mathbf{r}_k(\cdot)] - m_k). \end{aligned} \quad (23)$$

In order to determine the interval  $A_k$ , defined in (20), we derive the set of  $N_T - 1$  inequalities

$$\mathbb{E} \{ \Delta J[\mathbf{r}_j(\cdot)] \mid \Delta J[\mathbf{r}_k(\cdot)] \} > \Delta J[\mathbf{r}_k(\cdot)] \text{ for all } j \neq k. \quad (24)$$

The interval  $A_k$ , which is the range of values of  $\Delta J[\mathbf{r}_k(\cdot)]$  satisfying (24) is defined by

$$\max_{\mathbf{r}_j(\cdot) \in A^-} \frac{m_j - \tilde{\sigma}_{jk}}{1 - \frac{\tilde{\sigma}_{jk}}{m_k}} < \Delta J[\mathbf{r}_k(\cdot)] < \min_{\mathbf{r}_j(\cdot) \in A^+} \frac{m_j - \tilde{\sigma}_{jk}}{1 - \frac{\tilde{\sigma}_{jk}}{m_k}}, \quad (25)$$

where  $A^+$  is the set of all trajectories  $\mathbf{r}_j(\cdot)$  such that  $m_j < m_k$ ,  $\tilde{\sigma}_{jk} > m_j$ , and  $A^-$  is the set of all trajectories  $\mathbf{r}_j(\cdot)$  such that  $m_j > m_k$ ,  $\tilde{\sigma}_{jk} > m_k$ . Note that the supremum and infimum, over all continuous trajectories, of the leftmost and rightmost sides of (25), respectively, is  $-m_k$ . This means that as  $N_T$  increases and the trajectories  $\mathbf{x}(\cdot) + \mathbf{r}_i(\cdot)$  are sampled from  $\mathcal{C}_{[0,T]}$  according to their *a priori* distribution (3), the interval  $A_k$  narrows. Specifically, for any  $\delta, \tilde{\delta} > 0$  there is a sufficiently large  $N_T$  such that  $\Pr \{ A_k \not\subset (-m_k - \delta, -m_k + \delta) \} < \tilde{\delta}$ . Combining (19) and (21), we conclude that for all small  $\varepsilon > 0$  and every sufficiently small  $\delta$ , such that  $0 < \delta < \varepsilon$ , there is a sufficiently large  $N_T$  such that

$$\begin{aligned} \Pr \{ \mathbf{e}_{N_T}(\cdot) = \mathbf{r}_k(\cdot) \} &\asymp \Pr \{ -m_k - \delta < \Delta J[\mathbf{r}_k(\cdot)] < -m_k + \delta \} \\ &\asymp 2\delta \exp \left\{ -\frac{4 m_k^2}{8 \varepsilon m_k} \right\} \\ &\asymp 2\delta \exp \left\{ -\frac{1}{2\varepsilon} \int_0^T \left[ 4 \sin^2 \left( \frac{r_k}{2} \right) + |\mathbf{u}_k|^2 \right] dt \right\}. \end{aligned} \quad (26)$$

Based on the distribution of the estimation error  $\mathbf{e}_{N_T}(\cdot)$  in the case of a finite number of error trajectories  $N_T$  (26), the probability that  $\mathbf{e}_{N_T}(\cdot)$  is in any set  $\mathcal{A}$  in  $\mathcal{C}_{[0,T]}^N$  is

$$\begin{aligned} \Pr \{ \mathbf{e}_{N_T}(\cdot) \in \mathcal{A} \} &= \sum_{\mathbf{r}_k(\cdot) \in \mathcal{A}} \Pr \{ \mathbf{e}_{N_T}(\cdot) = \mathbf{r}_k(\cdot) \} \\ &\asymp \sum_{\mathbf{r}_k(\cdot) \in \mathcal{A}} 2\delta \exp \left\{ -\frac{1}{2\varepsilon} \int_0^T \left[ 4 \sin^2 \left( \frac{r_k}{2} \right) + |\mathbf{u}_k|^2 \right] dt \right\}. \end{aligned} \quad (27)$$

Applying Laplace's method for sums of exponentials with large parameter  $\frac{1}{\varepsilon}$  [3], we obtain the asymptotic expression

$$\Pr \{e_{N_T}(\cdot) \in \mathcal{A}\} \asymp \exp \left\{ -\frac{1}{2\varepsilon} \min_{\mathbf{r}_k(\cdot) \in \mathcal{A}} \int_0^T \left[ 4 \sin^2 \left( \frac{r_k}{2} \right) + |\mathbf{u}_k|^2 \right] dt \right\}. \quad (28)$$

In the limit  $\delta \rightarrow 0$  (and  $N_T \rightarrow \infty$ ), the probability that the optimal estimation error  $\mathbf{e}(\cdot)$  is in  $\mathcal{A}$  is found as

$$\Pr \{\mathbf{e}(\cdot) \in \mathcal{A}\} \asymp \exp \left\{ -\frac{1}{2\varepsilon} \inf_{\mathbf{r}(\cdot) \in \mathcal{A}} \int_0^T \left[ 4 \sin^2 \left( \frac{r}{2} \right) + |\mathbf{u}|^2 \right] dt \right\}, \quad (29)$$

subject to the equality constraint

$$\dot{\mathbf{r}} = \mathbf{A}\mathbf{r} + \mathbf{B}\mathbf{u}, \quad (30)$$

where the infimum in (29) is taken over all continuous trajectories  $\mathbf{r}(\cdot) \in \mathcal{A}$ . Note that the small  $\varepsilon$  approximation is taken before the limit  $\delta \rightarrow 0$ .

We turn now to the computation of the MTLT. First, we consider time intervals  $[0, T]$  much longer than the time constant of the system, so that most of the time the system is in steady state. It follows that the probability  $\Pr \{\text{slip in } (t_0, t_0 + \Delta t)\}$  is independent of  $t_0$  for  $t_0$  outside intervals of fixed length (the time constant of the system) at the endpoints 0 and  $T$ . Therefore, in view of the regularity of the pdf of the solution as a function of  $t$  and the independence of cycle slips in disjoint intervals (see above), for such  $t_0$

$$\begin{aligned} \Pr \{\text{slip in } (t_0, t_0 + 2\Delta t)\} &= \\ \Pr \{\text{slip in } (t_0, t_0 + \Delta t)\} + \Pr \{\text{slip in } (t_0 + \Delta t, t_0 + 2\Delta t)\} + o(\Delta t) &= \\ 2 \Pr \{\text{slip in } (t_0, t_0 + \Delta t)\} + o(\Delta t). \end{aligned} \quad (31)$$

Thus,  $\Pr \{\text{slip in } (t_0, t_0 + \Delta t)\}$  is nearly linear in  $\Delta t$ .

Next, we note that for fixed  $\Delta t$  (29) implies that the slip probability satisfies

$$\begin{aligned} &\Pr \{\text{slip in } (t_0, t_0 + \Delta t)\} \\ &= \Pr \{\mathbf{e}(\cdot) \in \mathcal{C}^N(t_0)\} \\ &\asymp \exp \left\{ -\frac{1}{2\varepsilon} \inf_{\mathbf{r}(\cdot) \in \mathcal{C}^N(t_0)} \int_0^T \left[ 4 \sin^2 \left( \frac{r}{2} \right) + |\mathbf{u}|^2 \right] dt \right\}, \end{aligned} \quad (32)$$

subject to the equality constraint (30). Equations (31) and (32) can be written together as

$$\begin{aligned} \Pr \{\text{slip in } (t_0, t_0 + \Delta t)\} &= \\ (\Delta t + o(\Delta t))\Omega(\varepsilon) \exp \left\{ -\inf_{\mathbf{r}(\cdot) \in \mathcal{C}^N(t_0)} \frac{1}{2\varepsilon} \int_0^T \left[ 4 \sin^2 \left( \frac{r}{2} \right) + |\mathbf{u}|^2 \right] dt \right\}, \end{aligned} \quad (33)$$



subject to the equality constraint (30), where  $\varepsilon \log \Omega(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

Equipped with the slip probability (33), we turn to the evaluation of the MTLL in the optimal smoother. We introduce a *renewal* (counting) process  $\{\mathcal{N}(t), t \geq 0\}$ , a nonnegative integer-valued stochastic process that counts the number successive cycle-slip events in the time interval  $(0, t]$  [11]. We assume that the time durations between consecutive slips are positive, independent, identically distributed random variables. Based on the above assumptions, we adopt the renewal formula [11]

$$\tau_{nc} = \frac{t}{\mathbb{E} \mathcal{N}(t)}, \quad (34)$$

where  $\tau_{nc}$  is the MTLL in the non-causal estimator. In the limit of  $t \rightarrow \infty$ , equation (34) gives

$$\tau_{nc} = \lim_{t \rightarrow \infty} \frac{t}{\mathbb{E} \mathcal{N}(t)} = \lim_{t \rightarrow \infty} \frac{1}{\mathbb{E} \dot{\mathcal{N}}(t)}. \quad (35)$$

Using the slip probability (33) we obtain

$$\begin{aligned} \mathbb{E} \dot{\mathcal{N}}(t) &= \lim_{\Delta t \rightarrow 0} \mathbb{E} \left\{ \frac{\mathcal{N}(t + \Delta t) - \mathcal{N}(t)}{\Delta t} \right\} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Pr \{ \text{slip in } (t_0, t_0 + \Delta t) \}}{\Delta t} \\ &\asymp \exp \left\{ - \inf_{\mathbf{r}(\cdot) \in \mathcal{C}^N(t_0)} \frac{1}{2\varepsilon} \int_0^T \left[ 4 \sin^2 \left( \frac{r}{2} \right) + |\mathbf{u}|^2 \right] dt \right\}. \end{aligned} \quad (36)$$

Substituting (36) in (35), we obtain the expression for the asymptotic MTLL,  $\tau_{nc}$ , in the optimal MNE estimator

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_{nc} = \inf_{\mathbf{e}(\cdot) \in \mathcal{C}^N(t_0)} \frac{1}{2} \int_0^T \left[ 4 \sin^2 \left( \frac{e}{2} \right) + |\boldsymbol{\xi}|^2 \right] dt, \quad (37)$$

subject to the equality constraint

$$\dot{\mathbf{e}} = \mathbf{A}\mathbf{e} + \mathbf{B}\boldsymbol{\xi}. \quad (38)$$

## 4 The MTLL in the smoother with standard phase models

We begin with the first order phase tracking system suggested by [27] and [18], in which the phase  $x(t)$  is modelled as a standard Brownian motion

$$\begin{aligned} \dot{x} &= \dot{w} \\ y &= \mathbf{h}(x) + \rho \dot{v}, \end{aligned} \quad (39)$$

where  $x(t), w(t)$  take values in  $\mathbb{R}^1$ . The system (39) is scaled to the form of (3), (4) with  $\mathbf{A} = 0, \mathbf{B} = 1$  and  $\varepsilon = \rho$ . A similar procedure is presented in [4]. Note that the CNR equals  $\rho^{-2}/2$  in the system (39).

In the first order case, the asymptotic expression for the MTLL in the smoother (37) becomes

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_{nc} = \inf_{e(\cdot) \in \mathcal{C}^1(t_0)} \frac{1}{2} \int_0^T \left[ 4 \sin^2 \left( \frac{e}{2} \right) + \dot{e}^2 \right] dt, \quad (40)$$

The variational problem on the right side of (40) is solved analytically using the Hamilton-Jacobi-Belman equation [12]. This leads to the asymptotic limit of the MTLL in the first order non-causal estimator as

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_{nc} = \lim_{\rho \rightarrow 0} \rho \log_e \tau_{nc} = 8. \quad (41)$$

In order to compare the MTLL in the optimal smoother (41) with that in the suboptimal PLL we construct the steady-state EKF corresponding to the model (39) [30].

$$\dot{\hat{x}} = \frac{\sigma}{\rho} (y_s \cos \hat{x} - y_c \sin \hat{x}). \quad (42)$$

The differential equation of the causal EKF estimation error  $e(t) = \hat{x}(t) - x(t)$  is scaled to the form [4]

$$\dot{e} = -\sin e + \sqrt{2\varepsilon} \dot{v}. \quad (43)$$

The asymptotic MTLL,  $\tau_c$ , in this simple analytical potential case is known to be [29]

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_c = \lim_{\rho \rightarrow 0} \rho \log_e \tau_c = 2. \quad (44)$$

Note that in first order estimators the MTLL in the non-causal estimator (41) is significantly longer than that in the causal estimator (44). This implies that the CNR values in the smoother are smaller than in the filter, but give a MTLL identical to that in the filter. Denoting by  $\varepsilon_c, \varepsilon_{nc}$  the values of  $\varepsilon$  in the filter and smoother, respectively, and requiring identical MTLLs, lead to

$$\tau_c = \tau_{nc} \Rightarrow \frac{2}{\varepsilon_c} = \frac{8}{\varepsilon_{nc}}. \quad (45)$$

Denoting by  $\text{CNR}_c[\text{dB}], \text{CNR}_{nc}[\text{dB}]$  the CNR in the filter and smoother, respectively, and using the simple relation between  $\varepsilon$  and the CNR, lead to

$$\text{CNR}_c[\text{dB}] - \text{CNR}_{nc}[\text{dB}] = 10 \cdot 2 \log_{10}(8/2) \approx 12\text{dB}. \quad (46)$$

Thus, there exists a 12dB performance gap in the MTLL between the estimators in terms of CNR. The MTLL in first order non-causal MNE smoother and causal EKF are given in Figures 2 and 3. The pre-exponential term in the plots is arbitrary.

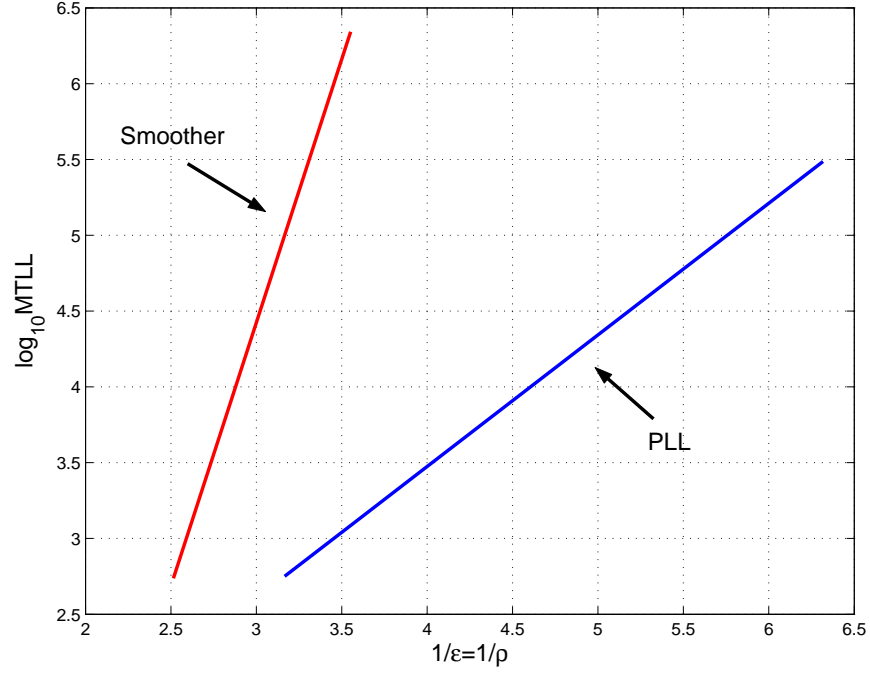


Figure 2: The MTLL in the first order optimal MNE estimator and the causal PLL as a function of  $1/\epsilon$ .

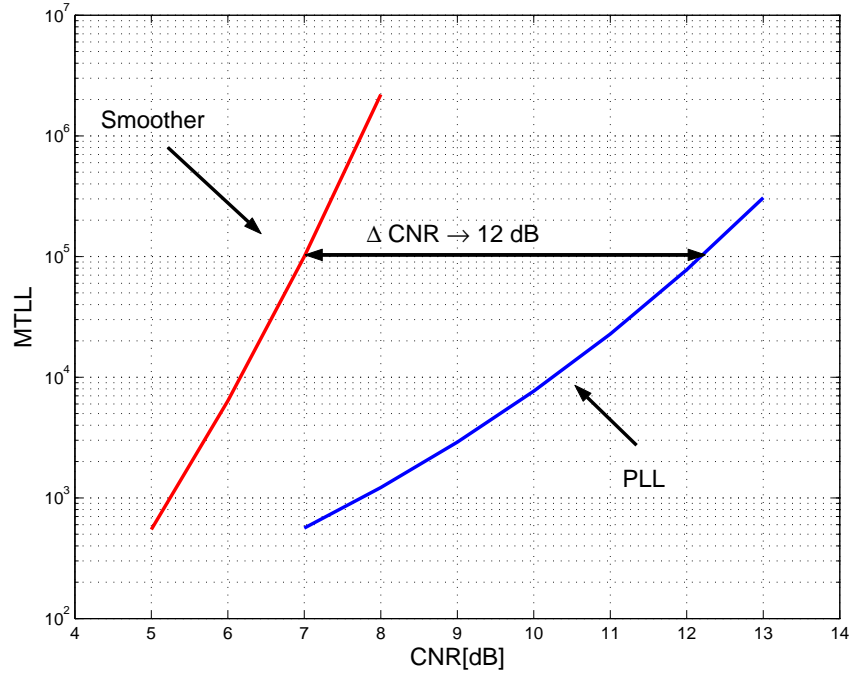


Figure 3: The MTLL in the first order optimal MNE estimator and the causal PLL as a function of the CNR.

Next we consider the more complex, and more realistic, case of a second order phase model [30], in which the phase is modelled as an integral over a Brownian motion. The signal model is

$$\dot{\mathbf{x}} = \mathbf{A}'\mathbf{x} + \mathbf{B}'\dot{\mathbf{w}}, \quad (47)$$

where

$$\mathbf{A}' = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B}' = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The measurements model is

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \rho \dot{\mathbf{v}}, \quad (48)$$

where  $\mathbf{x}(t)$  and  $\mathbf{w}(t)$  take values in  $\mathbb{R}^2$ . The system (47), (48) is scaled to the form of (3),(4) with  $\mathbf{A} = \mathbf{A}'$ ,  $\mathbf{B} = \mathbf{B}'$ , and  $\varepsilon = \rho^{3/2}$ . Note that in this case the CNR equals  $\rho^{-2}/2$ , as in the case of the first order system.

In the second order case, the asymptotic expression for the MTLL in the smoother (37) becomes

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_{nc} = \inf_{e(\cdot) \in \mathcal{C}^2(t_0)} \frac{1}{2} \int_0^T \left[ 4 \sin^2 \left( \frac{e}{2} \right) + \dot{e}^2 \right] dt, \quad (49)$$

variational problem on the right side of (37) as no analytical solution. An approximate solution is obtained numerically about the characteristics of the Hamilton-Jacobi-Belman equation [29, 5]. This leads to the asymptotic limit of the MTLL in the second order non-causal estimator as

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_{nc} = \lim_{\rho \rightarrow 0} \rho^{3/2} \log_e \tau_{nc} = 5. \quad (50)$$

We note that the EKF corresponding to the second order system (47), (48) has the error equations [30],[5]

$$\begin{aligned} \dot{e} &= \frac{1}{2}\varphi - \sin e - \sqrt{2\varepsilon'} \dot{v} \\ \dot{\varphi} &= -\sin e - \sqrt{2\varepsilon'} \dot{v} + \sqrt{2\varepsilon'} \dot{w}, \end{aligned} \quad (51)$$

where  $\varepsilon' = \varepsilon/\sqrt{2}$ . The "eikonal" equation [5] for the quasi-potential  $\Phi$  corresponding to (51) is

$$\left( \frac{1}{2}\varphi - \sin e \right) \Phi_e - \sin e \Phi_\varphi + \Phi_e^2 + 2\Phi_e \Phi_\varphi + 2\Phi_\varphi^2 = 0. \quad (52)$$

The function  $\Phi$  is evaluated numerically on the characteristics of (52) [5]. This procedure leads to the value of the quasi-potential at the unstable equilibrium point  $\Phi(\varphi = 0, e = \pi) = 0.6$ , so the asymptotic MTLL in the second order causal estimator is

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_c = \lim_{\rho \rightarrow 0} \rho^{3/2} \log_e \tau_c = 0.85. \quad (53)$$

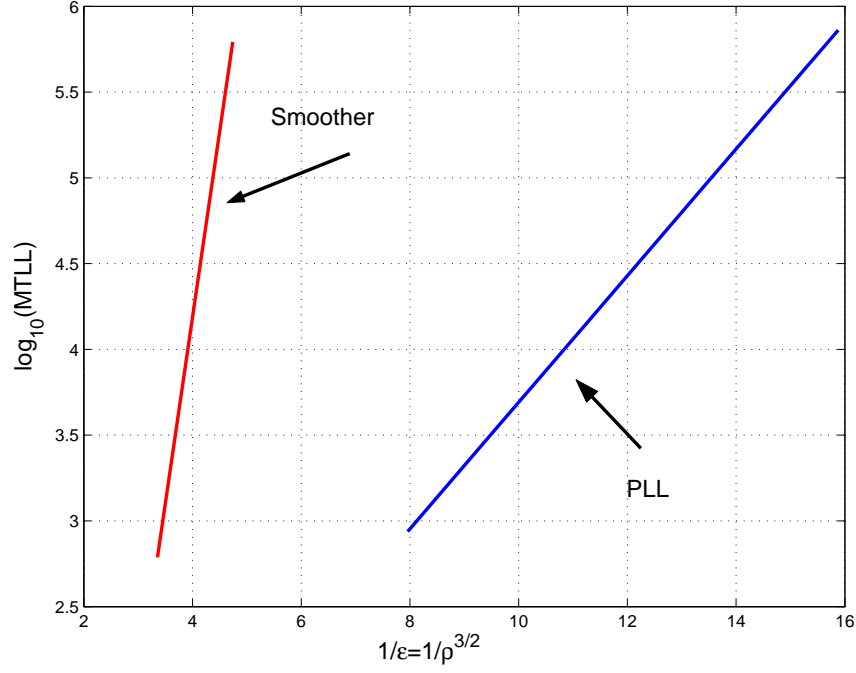


Figure 4: The MTLL in the second order optimal MNE estimator and the causal PLL as a function of  $1/\varepsilon$ .

Note that similarly to the first order case, the MTLL in the second order smoother (50) is significantly longer than that in the causal PLL (53). The CNR gap in this case is

$$\text{CNR}_c[\text{dB}] - \text{CNR}_{nc}[\text{dB}] = 40/3 \log_{10}(5/0.85) \approx 10.25\text{dB}. \quad (54)$$

The MTLL in the non-causal estimator and the causal EKF in the second order system are given in Figures 4 and 5. The pre-exponential term in the plots is arbitrary.

## 5 Discussion and conclusions

The significant advantage of the optimal smoother over the causal PLL defies intuition. It was customary to think that the MTLL advantage of the smoother is linked and proportional to the MSE advantage of the smoother [32]. We argue that the MSE in an estimator is not related to the MTLL. In order to demonstrate this idea we introduce the following error equation

$$\dot{e} = -2 \sin \frac{e}{2} + \sqrt{\varepsilon} \dot{\tilde{v}}. \quad (55)$$

The MSE in the linearized version of (55) is identical to that in linearized version of the system

$$\dot{e} = -\sin e + \sqrt{\varepsilon} \dot{\tilde{v}}. \quad (56)$$

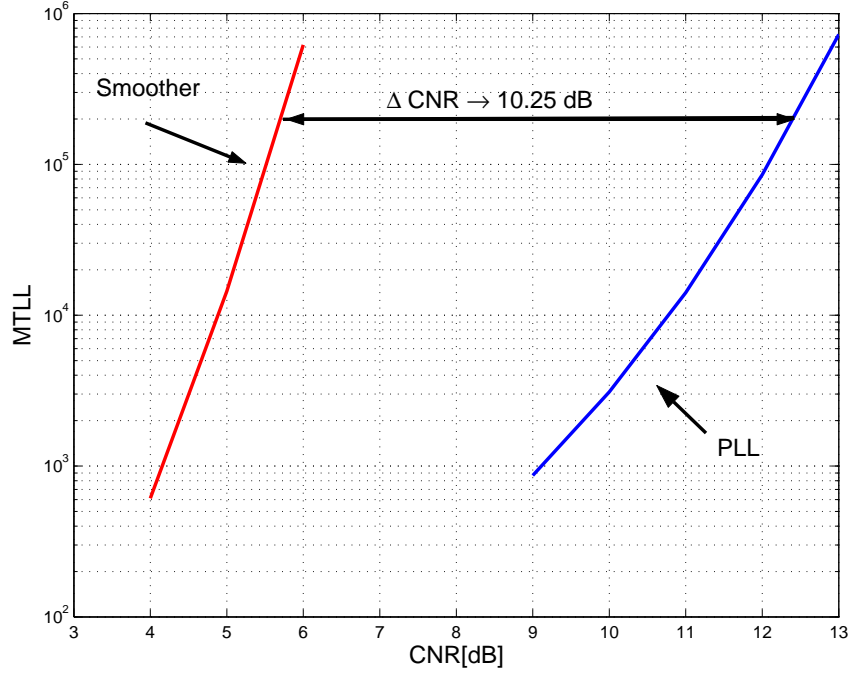


Figure 5: The MTLL in the second order optimal MNE estimator and the causal PLL as a function of CNR.

However, due to the difference in the potential barrier in the above systems, the asymptotical MTLL,  $\tau_1$ , in the system (55) satisfies

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_1 = \frac{1}{2} \inf_{\{e(0)=0, e(T')=2\pi\}} \int_0^{T'} \left( \dot{e} + 2 \sin \frac{e}{2} \right)^2 dt = 8, \quad (57)$$

while the MTLL,  $\tau_2$ , in the system (56) satisfies

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log_e \tau_2 = 4, \quad (58)$$

and is significantly shorter than  $\tau_1$ . Thus, we conclude that the MTLL advantage of the optimal smoother over the PLL is due to an entirely different functional and cannot be predicted by the MSE advantage of the smoother over the filter.

There is a fundamental mathematical difference between the causal and the non-causal cases. Both problems involve the minimization of a functional, similar to that of the Wentzell-Freidlin theory for causal systems. There is, however, a difference between the functionals in the two theories. While the functional in the causal case vanishes along the exiting trajectories from the boundary of the domain of attraction of the locked state to the next locked state [9], in the non-causal case the functional vanishes only at the locked states, so it has to be computed along the entire slip trajectory.

## References

- [1] Anderson, B.D.O. Fixed interval smoothing for nonlinear continuous time systems. *J. Info. Control*, 20:294–300, 1972.
- [2] Bellman, R., Kalba, R., and Middelton, D. Dynamic programming, sequential estimation and sequential detection processes. *Proc. Nat. Acad. Sci. U.S.A.*, 47:338341, 1961.
- [3] Bender, C.M., and Orszag, S.A. *Advanced Mathematical Methods for Scientists and Engineers*. Springer, New York, 1999.
- [4] Bobrovsky, B.Z., and Schuss, Z. Singular perturbation in filtering theory. *IFAC Workshop on singular perturbations in optimal control*, pages 1439–1447, June 1978.
- [5] Bobrovsky, B.Z., and Schuss, Z. A singular perturbation method for the computation of the mean first passage time in a non linear filter. *SIAM*, 42(1):174–187, Feb. 1982.
- [6] Bryson, A.B., and Ho, Y.C. *Applied Optimal Control*. John Wiley, New York, 1975.
- [7] Bucklew, J.A. *Large Deviation Techniques in Decision, Simulation and Estimation*. Wiley, New York, 1990.
- [8] Dembo, A., and Zeitouni, O. *Large Deviations Techniques and Applications*. Jones and Bartlett, 1993.
- [9] Freidlin, M.A., and Wentzell, A.D. *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York, 1984.
- [10] Kailath, T., and Frost, P. An innovation approach to least-squares estimation part II: Linear smoothing in additive white noise. *IEEE Trans. Auto. Control*, AC-13:655–660, 1968.
- [11] Karlin S., and Taylor, H.M. *A First Course in Stochastic Processes*. Academic Press, 1975.
- [12] Kirk D.E. *Optimal Control Theory - an Introduction*. Prentice-Hall, Inc., 1970.
- [13] Kushner, H.J. Dynamical equations for optimal nonlinear filtering. *J. Diff. Equations*, 2:179–190, 1967.
- [14] Lee, R.C.K. *Optimal Estimation Identification and Control*. M.I.T. Press, 1964.
- [15] Leondes, C.T., Peller, J.B., and Stear, E.B. Nonlinear smoothing theory. *IEEE Trans. Sys. Sci. Cyb.*, SSC-6:63–71, 1970.

- [16] Lindsey, W.C. *Synchronization Systems in Communication and Control*. Prentice-Hall, Englewood Cliffs, 1972.
- [17] Liptser, R.S., and Shiryayev, A.N. *Statistics of Random Processes*, volume I,II. Springer-Verlag, New York, 1977.
- [18] Macchi, O., and Scharf, L.L. A dynamic programming algorithm for phase estimation and data decoding on random phase channels. *IEEE Trans. Info. Theory*, IT-27(5):585–595, Sept. 1981.
- [19] Meditch, J.S. A survey of data smoothing for linear and nonlinear dynamic systems. *Automatica*, 9:151–160, 1973.
- [20] Proakis, J.G. *Digital Communications*. McGraw-Hill, 4th edition, 2001.
- [21] Rauch, H.E. Linear estimation of sampled stochastic processes with random parameters. Technical Report 2108, Stanford Electronics Laboratory, Stanford University, California, 1962.
- [22] Rauch, H.E. Solutions to the linear smoothing problem. *IEEE Tans. Auto. Control.*, AC-8:371–372, 1963.
- [23] Rauch, H.E., Tung, F., and Steibel, C.T. Maximum likelihood estimates of linear dynamic systems. *AIAA J.*, 3:1445–1450, 1965.
- [24] Ryter, D., and Meyr, H. Theory of phase tracking systems of arbitrary order. *IEEE Trans. Info. Theory*, IT-24:1–7, 1978.
- [25] Sage, A.P. Maximum a posteriori filtering and smoothing algorithms. *Int. J. Control*, 11:171–183, 1970.
- [26] Sage, A.P., and Ewing, W.S. On filtering and smoothing algorithms for nonlinear state estimation. *Int. J. Control*, 11:1–18, 1970.
- [27] Scharf, L.L., Cox, D.D., and Masreleiz, C.J. Modulo  $2\pi$  phase sequence estimation. *IEEE Trans. Info. Theory*, IT-26(5):615–620, Sept. 1980.
- [28] Schuss, Z. Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM*, Rev 22:119–155, 1980.
- [29] Schuss, Z. *Theory and Applications of Stochastic Differential Equations*. John Wiley, New York, 1980.
- [30] Snyder, D.L. *The State Variable Approach to Continuous Estimation*. The M.I.T Press, 1969.
- [31] Tausworth, R. Cycle slipping in phase locked loops. *IEEE Trans. Comm.*, COM-15:417–421, 1967.



- [32] Van Trees, H.L. *Detection, Estimation and Modulation Theory*, volume II. John Wiley, New York, 1970.
- [33] Viterbi, A.J. *Principles of Coherent Communication*. McGraw-Hill, New York, 1966.
- [34] Viterbi, A.J., and Omura, J.K. *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.
- [35] Weaver, C.S. Estimating the output of a linear discrete system with Gaussian input. *IEEE Trans. Auto. Control*, AC-8:372–374, 1963.